Eyes, Screen, and Hands: Investigating Spatial Manipulation in Peripersonal Space

ABSTRACT
Advances in virtual and augmented reality (VR/AR) technology are re-shaping our way of interacting with the virtual world. However, current VR/AR methodologies still fall short of being the default in how humans approach and interact with their virtual surroundings. Our work draws from ecological psychology and action-specific perception to re-think spatial interactions for object manipulation in the context of peripersonal space, space immediately surrounding the body. In this paper, we focus on the relative spatial configuration of the user’s visual field (eyes), the location of visual feedback (screen), and the space where objects are manipulated (hands). We specifically study and compare two configurations, namely, “eyes→hand→screen” and “eyes→screen→hand” in spatial manipulation tasks. For this, we implemented a desktop-based non-immersive VR system powered by small-space motion-tracking for users to manipulate small scale 3D objects through hand-held controllers. Our qualitative and quantitative results highlight a need to explore the domain of spatial interactions for high precision tasks in the user’s peripersonal space.

INTRODUCTION
Manual tasks such as sculpting or assembly of parts involve close, careful, and precise handling of tools and work-pieces, often with both hands. In such tasks, the space where the action takes place is co-located with the space where the action is perceived. Gibson’s seminal adage that “perception is for action” informs our research [29]. The corollary of this concept for design of interaction is that action should be designed to match the powers of human perception. Human binocular perception, fine motoric proprioception, and sensitivity of fingertip tactile perception are paired with the fine motor control to enable humans to be the “tool-making beings” thereby extending human ability far beyond the limitations of our biology. The conceptual prototype that captures the essence of this extended ability is the clock-maker’s work-space (also referred to as CMWS throughout this paper). The space in which this high precision interaction is enacted is known as the peripersonal space (the space immediately surrounding the body) as opposed to the extrapersonal space (the space farther away from the body) [40]. Our goal in this paper is to address some fundamental methodological gaps in current HCI practices for spatial object manipulation by investigating and highlighting the role of peripersonal spaces.

Spatial interaction is a well-established research area that has gained more popularity particularly with the commodification of augmented and virtual reality (AR/VR) systems. Consequently, the domain of spatial interactions in AR/VR environments has advanced significantly [9, 2]. In principle, these developments are aligned with the embodied interactions viewpoint wherein the intention is to incorporate bodily practice into interactions with virtual artifacts such that users perceive the artifact as an extension of themselves; they act through it rather than on it [43]. However, the central goal of integrating sensory perception and motor skills towards controlled spatial manipulation of virtual objects is far removed from what is both possible and needed for future spatial user interfaces. The fundamental problem is that the distinction between peripersonal and extrapersonal spaces is currently missing from the AR/VR interaction design toolbox.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-6708-0/20/04…$15.00.

DOI: https://doi.org/10.1145/3313831.XXXXXX
that affect the precision of user action during a precise spatial manipulation task. As far as current spatial interaction research is concerned, how these factors affect spatial manipulation for peripersonal space in comparison to extrapersonal space is an open question. For instance, we do not currently know if bi-manual (using both hands) interactions are always faster and more accurate across peripersonal and extrapersonal spaces. Similarly, we also do not know how peripersonal and extrapersonal spaces compare with respect to bi-manual manipulation. Most importantly, there is currently limited understanding of how these two spaces affect the physical and mental load on the user. In this paper, our primary goal is to systematically and critically study spatial manipulation from the perspective of ecological psychology to address this question.

Our work specifically focuses on the relative spatial configuration of the user’s visual field (eyes), the location of visual feedback (screen), and the space where objects are manipulated (hands). We hypothesize that a difference of configuration from “eyes→hand→screen” (that is currently the norm) to “eyes→screen→hand” (i.e. behind-the-screen interactions) will help us better understand the action-specific perception in spatial manipulations. The inspiration for our work comes from some seminal systems such as HoloDesk [35], SpaceTop [50], and MixFab [74] that underscore the importance of spatial interactions in peripersonal space. However, these and several related works are primarily focused on specific applications such as digital and physical prototyping. There is currently no in-depth analysis of precise bi-manual manipulation of small objects in the user’s intimate and personal space.

Contributions:
Our work complements the existing literature by providing a systematically developed and empirically verified interface and evaluation methodology for peripersonal spatial object manipulation. While some of the findings presented here align with those already known in psychology and motor behavior, our work reveals some new insights that may be useful in the design of physical experiences. From the perspective of neuropsychology, peripersonal and extrapersonal spaces are defined by our brains with our body as the reference — peripersonal space is “centered on body parts (i.e., hand-centered, head-centered, and trunk-centered)” and “is for the interaction with objects and people in the space around us” [21]. In fact, peripersonal space is not a “single, distance-based, in-or-out zone” [12]. Bufacchi et al. [12] show evidence that indicates a more fuzzy, action-dependent, context-dependent nature of peripersonal space that can have multiple components (say a union of volumes around the head, torso, and hands) and also changes due to factors other than proximity.

Longo and Lourenco [53] and Witt et al. [81, 80] provided early evidence to this fact and noted that the perception of what comprises “near space” is flexible and changes with usage of tools. Specifically, tools increase the mental range of near space (at least within the range of around 120 cm). As a result, the fundamental behavior in object manipulation can change drastically based on the space where objects are located [62]. Davoli [18] noted: “By shifting an object from extrapersonal space to peripersonal space, the object may appear closer while remaining in the same physical location, an effect that would have implications for how one chooses to perform visually controlled tasks.” This is a critical observation that needs to be emphatically incorporated in the design of spatial manipulations. In this regard we note that the design of precise spatial manipulations should: (a) position the objects and the action at the intersection of the three peripersonal spaces around the head, hands, and torso; and (b) integrate the use of hand-held tools as a means to enabling economic actions [62].

• First, we provide a concrete implementation of the principles of ecological psychology and action-specific perception by developing a physical embodiment of the clockmaker’s work-space (CMWS). We implemented CMWS re-configuration a motion capture system for close-range motion tracking and implemented a tablet-based virtual environment that is controlled through a virtual hand-held controller (e.g. Wii Remote). The key distinction between our prototype interfaces with respect to default VR/AR systems is that our prototype, while non-immersive, provides exclusively virtual visual feedback (as in VR). However, it is spatially configured as a through-the-screen (“eyes→screen→hand”) interfaces. While such interfaces have been developed prior to this work, our main contribution lies in bridging the theory of precise motor control with the design of spatial user interfaces followed by an in-depth study of precise bi-manual spatial manipulation.

• Second, our specific implementation allowed us to conduct an in-depth two-way comparison of spatial object manipulation in peripersonal space with our designed interface using both uni-manual (using one hand) and bi-manual interactions. For our comparison, we developed a display-based conventional non-immersive VR (CVR) system that mirrors the “eyes→screen→hand” configuration. While we discovered that neither configuration (CMWS and CVR) was unilaterally better than the other for all factors (object size, object shape, uni-manual manipulation, and bi-manual manipulation), the main advantage of CMWS was clearly in alignment with users’ action-specific perception as well as their experience. Furthermore, it also revealed that CMWS lends itself to bi-manual manipulation while CVR leads the user to perform interactions at arms-length thereby forcing the user to shift to uni-manual manipulation due to fatigue.

Finally, we consolidate the challenges faced in the development of our interface with our experimental findings to provide a set of design guidelines for peripersonal spatial object manipulation.

BACKGROUND
Action & Perception
The notion of peripersonal space is tightly linked to Gibson’s position of ecological psychology [29], which gave rise to the notion of perceived affordance, a concept we use abundantly in the design of physical experiences. From the perspective of neuropsychology, peripersonal and extrapersonal spaces are defined by our brains with our body as the reference — peripersonal space is “centered on body parts (i.e., hand-centered, head-centered, and trunk-centered)” and “is for the interaction with objects and people in the space around us” [21]. In fact, peripersonal space is not a “single, distance-based, in-or-out zone” [12]. Bufacchi et al. [12] show evidence that indicates a more fuzzy, action-dependent, context-dependent nature of peripersonal space that can have multiple components (say a union of volumes around the head, torso, and hands) and also changes due to factors other than proximity.
Proxemics in Spatial User Interfaces

Traditional spatial user interfaces in AR/VR systems integrate medium to large interaction volume for 3D object manipulation tasks [9, 27]. On the other hand, interactions in peripersonal space have primarily been contextualized for studying social behavior of virtual avatar of humans in VR technologies for enhancing user engagement [56, 44, 22, 84], proximity to interactive display setups [8, 54, 55, 59, 52, 41], or mechanisms for controlling software elements remotely [48]. In almost all cases, the focus is on affecting feedback on the visual display once the user is in the proximity of the display [8, 55]. In other words, the context for a given action is already pre-determined in terms of the location of the display that is often a large device (at least from a peripersonal space perspective) and is also far away from the user. In other words, the user ends up making a conscious effort to first reach the environment (both physically and mentally) as pre-cursor to manipulation. As a result, there are no definitive guidelines to guide the design of small interaction volume for high precision tasks.

In fact, there are only a few works that shed light on the fundamental interactive aspects of peripersonal spaces. For example, Lee et al. [49] introduced adaptive gaze control in peripersonal spaces. Other works such as HoloDesk [35], SpaceTop [50], and MixFab [74], that capture peripersonal interactions are by and large application-oriented. While there are in-depth studies on sketching in VR applications [3] and docking [72], they are still representative of the large-screen distant interaction approach to spatial manipulation. Our intention is to address this methodological gap through a systematic investigation of spatial manipulation in peripersonal space.

Proprioceptive Feedback

Proprioceptive and kinesthetic control are inherent to humans in any physical interaction and are the key to design processes that involve human action and perception [26]. The lack of kinesthetic control severely impedes spatial manipulation. What specifically is relevant to peripersonal spaces is ability to perform finger-level manipulations. This view is echoed in several works contextualizing proprioception for 3D user interfaces, specifically for pointing and selection of interface elements in virtual environments [17, 83]. Proprioception at an egocentric distance around the periphery of human body reduces dependency on visual feedback, generally for local and distant mid-air interaction spaces. Works by Plaumann [60] and Popvici [61] put this view into practice by developing displayless IoT device interfaces by spatially placing virtual controls and shortcuts in user’s personal, peripersonal, and extrapersonal spaces. The visual dependency in virtual environments is attributed to distinct motor and visual spaces for spatial interactions [2]. On similar lines, work by DeBoeck et al. [19] also demonstrates that interactions in close proximity to the body improves kinesthetic control by exploiting proprioception. The key observation relevant for our work is that while proprioception enhances manipulative precision close to the body, the same is not true for for distal interactions, as shown previously in the case of target selection [28].

Bi-manual Action in Spatial Interactions

Work by Hinckley et al. [39] studied cooperative bi-manual interactions for virtual manipulation and provides a strong evidence for augmenting hand-eye coordination through the use of two hands in conjunction with haptics feedback. Two-handed interactions in coordinated tasks have been shown to increase cognitive engagement of the user [51, 64, 63] and efficiency of 3D object assembly [31]. Several works explore the advantages of bi-manual spatial interactions [15, 68, 38, 7, 6, 82, 67, 66] for object selection [70] and manipulation (rotation, translation, scaling) of 3D objects. Brandl et al. [10] explore the combination of two-handed interactions with pen and multi-touch inputs on a surface. Regardless of the wealth of literature, we believe that much is to be discovered regarding bi-manual interactions, when seen in the context of proprioception in peripersonal spaces. For this, one of our primary goals is to observe the differences between uni-manual and bi-manual actions in the context of precise manipulation. Here, the condition that the only visual feedback to the user is through a virtual display instead of the direct observation of the hands interacting with the object in physical space to elicit a co-located visual and motor space, also, to reduce occlusion of the visual rendering [4, 5, 25].

DESIGNING THE CLOCK-MAKER’S WORK-SPACE

In this paper, our goal is to investigate high-precision bi-manual tasks. In the following sections, we discuss and elaborate on the rationale and methodology for designing the interaction spaces for 3D object manipulation in AR/VR systems while preserving the user’s physical peripersonal space. There are three primary factors that guided the design and implementation of the clock-maker’s work-space for peripersonal spatial manipulation.

• Anatomy: Typically, peripersonal space varies with the user’s body structure, further classified based on the body-part in action – hands, face, and trunk [65, 21]. In this paper, we focus on the peripersonal space defined by the upper limbs, specifically the hand, i.e. the perihand space, mainly to observe and analyze comfortable interaction distance for precision tasks in AR/VR systems and its relation to the anatomical peripersonal space.

• Controller size & shape: In spatial user interfaces, the upper limbs regulate the range of motion for object manipulation interactions. Also, an additional degree of control is facilitated by the type of grasp for holding objects during spatial manipulation [14, 11]. Different object sizes are held with a grip which typically vary from two to five fingers, regulating the kinematic constraint of the user’s hand motion [45, 13, 16, 58, 57]. In addition to the object size, its geometry also affects the grasp anatomy of how the object is being held. Therefore, we tested different commercially available controllers and settled for a Wii Remote that allowed for a using a cylindrical grasp, in turn providing a power grip for spatial manipulation [24, 47]. The controller was simply intended to be a physical proxy and wasn’t wirelessly connected to the setup.
Setup

User
Tablet PC
Intimate 
Space
Personal 
Space
Input Mode

a. Clock-Maker’s Work-Space
b. Conventional Non-Immersive VR

User
Input Mode

6 DoF Motion 
Capture System

Input Mode

Figure 1. (a) Motion capture study setup for Clock-Maker’s Work-Space (CMWS), (b) Motion capture study setup for display-based Conventional non-immersive VR (CVR) system, (c) bi-manual, and (d) uni-manual controller input modes.

• Manipulation Distance & Working Volume: There are two geometric aspects of the work-space that demand attention. First, distance from the torso is one of the key factors that help define the peripersonal space. The second factor is the volume of space defined by the limits of the manipulation task. Here, we invoke action field theory proposed by Bufacchi and Ianetti [12] to determine how far the manipulation should be. An important factor to consider here is that the manipulation is assumed to be performed with a hand-held object. Based on this, Bufacchi and Ianetti prescribe a distance within the range of 45 — 60 centimeters (1.5 — 2 feet). As for the working volume, we follow an iterative approach starting from a standard table-top dimensions. The main challenge here is to determine a reasonably small working volume, suitable for precise manipulations and still maintain robust object tracking.

CMWS Physical Setup

Motion Tracking

Our study setups (Fig. 1(a),(b)) include a tablet computer running a digital application, and a 6 degree-of-freedom (DoF) motion capture camera system for recording the position and orientation of the hand-held controllers used for 3D object manipulation. Through preliminary experiments with the motion capture setup and calibration accuracy, we designed an interaction volume measuring 2 ft x 2 ft x 1.5 ft that was able to efficiently track the hand-held controllers held by the users using a cylindrical grasp in a power grip stance. (Figs. 2(a), (b)). Generally, single-camera setup motion tracking devices [75, 78, 76] have small interaction volumes for tracking hand movements. Given our need for high-precision tracking, robust, sensitive, and accurate mapping of user movements is efficiently facilitated by 6 DoF motion capture cameras. The idea is to capture and record a small spatial interaction volume allowing high-precision 3D object manipulation through hand-to-object coupling [46].

Interaction Space

In the context of peripersonal space, the interaction proximity of a user typically varies between their intimate to personal space (1.5 ft to 4 ft), within the user’s arm length [79, 30, 33]. However, from the point of view of current AR/VR systems (Fig. 1(b)) users generally interact at the extrapersonal distance (beyond their arm’s length) for any spatial task, which contradicts the fundamental approach for high-precision manipulation tasks in the physical world. For this, we designed the motion capture setup to track the interaction space behind the display (Figs. 1(a)). This design decision was based on two primary rationale: (a) co-locating the virtual and physical (motor) peripersonal space for high-precision tasks; and (b) reduce occlusion caused by placing the user’s hands in front of the screen [4, 5, 25]. Our intent is to enable a degree of control similar to physical interactions for the user to perform precise spatial manipulation actions in the virtual environment, as corroborated for most small to medium space interaction volumes for high precision tasks [23]. In peripersonal space, the visual perception of object size dominates over actual object size. This has already been recorded in psychological studies on action-specific perception [81, 80]. The intent is to statistically identify the most commonly preferred viewing and interaction distance from the user providing the most accurate response to user movement.

Implementation

Our experimental setup (Fig. 1(a),(b)) is comprised of a Microsoft Surface Book laptop computer with an Intel Core i7-6600U CPU (3.4GHz), 16GB of GDDR5 RAM, and an NVIDIA GeForce GTX 965M graphics card, running 64-bit Windows 10 Professional Operating System. For motion capture, we used eight Optitrack Flex 13 cameras with a field of view of 56° and a refresh rate of 120 Hz. Our user evaluation...
We designed a between-subject experimental study to evaluate high-precision 3D object manipulation tasks in order to compare: (1) Clock-Maker’s Work-Space (CMWS) with display-based Conventional non-immersive VR (CVR) system in terms of user precision and (2) uni-manual and bi-manual object manipulation. We specifically make these comparisons across virtual objects of two different types of shapes (cylindrical and trapezoidal) and two different size groups. Therefore, for each shape and each size group, we independently compare: (1) uni-manual and bi-manual manipulation for CMWS, (2) uni-manual and bi-manual manipulation for CVR, and (3) CMWS and CVR for bi-manual manipulation.

EXPERIMENT

We designed a between-subject experimental study to evaluate high-precision 3D object manipulation tasks in order to compare: (1) Clock-Maker’s Work-Space (CMWS) with display-based Conventional non-immersive VR (CVR) system in terms of user precision and (2) uni-manual and bi-manual object manipulation. We specifically make these comparisons across virtual objects of two different types of shapes (cylindrical and trapezoidal) and two different size groups. Therefore, for each shape and each size group, we independently compare: (1) uni-manual and bi-manual manipulation for CMWS, (2) uni-manual and bi-manual manipulation for CVR, and (3) CMWS and CVR for bi-manual manipulation.

Participants

The participant group consisted a mix of 22 (12 female, 10 male) graduate and undergraduate students (18-30 years old) from engineering, architecture, and visualization majors. Of these, 9 participants had prior experience with VR/AR systems, and 4 of them had experienced mid-air gesture and interactions either through gaming consoles, mixed-reality headsets, or prior human-subject studies related to VR. We grouped the participants into two equal groups on the basis of spatial interaction approaches, (1) CMWS and (2) CVR (Figs. 1(a), (b)); so as to evaluate high-precision 3D object manipulation tasks in the user’s peripersonal space.

Evaluation Tasks

Based on our experimental goals, we designed the following tasks to compare CMWS and CVR for high-precision 3D object manipulation:

Task Description:
All user trials involved a general shape-assembly (Fig. 3) based task that is commonly used for evaluating 6DoF control [31] and can be generalized for CAD systems requiring high precision bi-manual control [70]. The participants were shown a pair of virtual 3D meshes on the study interface akin to a peg and a hole, also, each peg-hole 3D mesh-pair had two size variants - `large` and `small`. The goal of this task was for the participants to spatially manipulate the position and orientation, and dock the peg inside the hole, aligning the longitudinal axes of these shape-pairs. In order to have as-close-as possible correlation between the physical spatial action and its visual representation on the tablet display, a base dimension was chosen heuristically for the smaller versions of both cylinder and the trapezoid. The larger versions of the respective shapes were designed to have the same length as their smaller counterparts, but with a 20% increase in based shape dimensions — diameter for the cylinder and increase in the height for the trapezoid. The docking task was meant to allow for us to quantitatively assess user performance in terms of shape orientation and completion time across interaction approaches — (uni-manual and bi-manual) for each configuration, and subsequently across the configurations — (CMWS and CVR).

We chose cylinder and trapezoidal prism (for brevity, we use the term “trapezoid”) as our target geometries for the study task so as to control: (a) the rotational symmetry of each shape during spatial orientation; and (b) alignment of shape edges during docking. The cylinder has a relatively smoother surface and edge geometry compared to the trapezoid, and was hypothesized to be docked with more ease. As a result, the trapezoid with relatively sharper edge geometry and lack of rotational symmetry would take more time to dock in its corresponding hole. We did not have a strict time limit for the tasks, however, each trial was controlled to be completed between 8 and 12 seconds.

With this general approach, we conducted the following between-subjects user evaluation studies based on input dexterity across the two study configurations.

• Bi-manual Docking: In this task the user was asked to bi-manually control the peg-hole mesh-pair shown on the study interface (Fig. 3) using hand-held controllers (Fig. 2(a)). As mentioned in Sec. 3.1, task-based spatial interactions were performed either at a distance typically within the intimate space (CMWS) of the user or beyond the intimate space (CVR) depending on the study configuration (Fig. 1(a),(b)). Based on the user’s handedness, each hand was mapped either to control the peg or the hole, and subsequently dock them together using bi-manual co-ordination. Further, the user was asked to provide qualitative feedback on their experience with the assembly-based docking tasks using the NASA task-load index [34].

• Uni-manual Docking: In contrast to the bi-manual task, this approach allows for the user to control only the peg uni-manually while the hole was fixed in virtual space at an elevation of 45°. The elevation was decided by preliminary experiments to minimize wrist-based constraints, thus, providing a comfortable uni-manual docking experience for both setup configurations — CMWS and CVR. Similar to the bi-manual approach, `small` and `large` shape sizes were provided for the uni-manual docking tasks, and the qualitative user feedback was collected using the NASA task-load index [34] on completion of the tasks.
Procedure

The experiment took approximately 45 minutes per user for each configuration (CMWS and CVR) to compare across uni-manual and bi-manual manipulation. Each session started with the general introduction of the motion capture system and the study interface, familiarizing the users with our proposed way of interacting behind or before the tablet display depending on the study configuration. In order to reduce tracking noise, users were asked to remove any jewellery and accessories that could cause interference with the motion tracking. This was followed by an initial demographic questionnaire highlighting any prior VR and spatial manipulation experiences. The experiment subsequently consisted of the following tasks:

Practice: Once familiarized with the motion capture system, the participants practiced spatial manipulation by manipulating a set of sample shapes using the hand-held controllers for the initial 5 minutes of the study. We ensured that adequate practice was provided in terms of spatial positioning, orientation, and motor-visual correlation for the docking assembly-tasks before starting with the actual study trials.

Trials: Each trial had an identical goal of docking a 3D peg into a 3D hole represented by virtual 3D shapes in the study interface. These virtual shapes are manipulated by hand-held controllers for multiple trials randomized using the Latin Square [77] approach across shape types — cylinder and trapezoid, also, their individual size variants — small and large.

Each setup configuration has two manipulation approach — uni-manual and bi-manual, experimented by the participants in their respective study groups — CMWS and CVR. Each participant performed 12 trials per shape per size variant (e.g. small Cylinder, large Trapezoid, etc), performing a total of 48 trials for all shapes and sizes across both manipulation approaches. Subsequently, total of 1056 trials were recorded for 22 participants enrolled for the evaluation user studies with 528 trials per study group.

Data & Metrics

For each trial performed by a participant, we recorded the raw event log containing a time-stamped controller trajectory where each trajectory point included the 3D position of the hand-held controller representing both the peg and the hole; the orientation of the entire local coordinate frame of the controllers. We additionally recorded the final orientation of the peg with respect to the hole, the user feedback provided in our online questionnaires, and the video of the participant performing the tasks.

We chose one-half of the magnitude of the cross product (Fig. 4) for measuring angular precision. The primary reason being that cross product represents signed area of a parallelogram formed by the normals of the base and target shapes. Thus, half the magnitude simply represents the minimum manual work needed to close the angular gap between the two shape normals. In addition, it provides a theoretically sound reasoning for the inverse relationship cause by the dot product. It can be derived from the identity: $\cos^2 \theta + \sin^2 \theta = 1$. The alignment error is thus computed as:

$$E_{\text{deviation}} = \frac{1}{2} \sin \theta \quad (1)$$

Here, $\theta$ is the angle between the orientation vectors of the base and target shape (that, in our case, is the $z$-axis).

RESULTS

In the following sections, we report on the statistical analysis on comparison of the two study configurations — Clock-Maker’s Work-Space (CMWS) and non-immersive Conventional VR (CVR) system. Further, we discuss the main insights gained from our data collection, observation, and user-feedback from the trials performed by all participants. First we present a pair-wise comparison of the uni-manual and bi-manual manipulation approaches for each of the two configurations (section 5.1, section 5.2). Subsequently, we shift our focus to the comparing two study configurations (section 5.3).

CVR: Pairwise Uni-manual vs. Bi-manual Manipulation

In this sub-section we evaluate the uni-manual and bi-manual manipulation approach for the non-immersive screen-based Conventional VR system. In this study, the spatial actions performed by the participants were observed to be away from the user’s body in the extrapersonal space.

User Performance: Alignment Error

In order to evaluate the final orientation between the peg and the hole docked by the participants, we make the following hypotheses:
Null($H_o$): There is no significant difference in the alignment errors for the docking-assembly task in uni-manual and bi-manual input modes.

Alternate($H_a$): There is a significant difference between the alignment errors for the docking-assembly task in uni-manual and bi-manual input modes.

The data samples collected for the uni-manual and bi-manual docking tasks (Figure 5) are evaluated for normal distribution using the Shapiro-Wilk test and normality is disregarded for p-values < 0.05. Since none of the data samples were found to be normally distributed, we perform statistical analysis using the non-parametric Kruskal-Wallis test for hypothesis testing based on the aforementioned null($H_o$) and alternate($H_a$) hypotheses. We observe a statistically significant difference in the alignment errors for the small and large versions of the cylindrical shape with p-values < 0.01, favoring the bi-manual manipulation. The median values for the bi-manual input mode were found to be 0.13 and 0.14 compared to 0.22 and 0.2 in the uni-manual mode for the small and large cylindrical shapes respectively. In case of the trapezoid, the median errors were found to be close with no significant difference for the uni-manual and bi-manual input modes having 0.7 for both the smaller and larger trapezoids with p-values of 0.45 and 0.37 respectively across the uni-manual and bi-manual input modes.

**User Performance: Docking-Task Completion Time**

In order to evaluate the time taken by the participants to completely dock the peg into the hole, we make the following hypotheses:

Null($H_o$): There is no significant difference in the time taken to completely dock the peg and the hole assembly in uni-manual and bi-manual input modes.

Alternate($H_a$): There is a significant difference in the time taken to completely dock the peg and the hole assembly in uni-manual and bi-manual input modes.

The data samples collected for the uni-manual and bi-manual docking tasks (Figure 6) are evaluated for normal distribution using the Shapiro-Wilk test and normality is disregarded for p-values < 0.05. Similar to the error metric, all data samples for task completion time were not found to be normally distributed. Therefore, we perform statistical analysis using the non-parametric Kruskal-Wallis test for hypothesis testing based on the aforementioned null($H_o$) and alternate($H_a$) hypotheses. No statistically significant difference was observed across both the 3D shapes and their corresponding size variants. The median completion times for the small cylinder were 7.1 seconds and 6.9 seconds for the uni-manual and bi-manual approaches respectively, whereas, 6.6 seconds and 5.7 seconds respectively for the large cylinder across both manipulation approaches. In case of the trapezoid, the median completion times for the small trapezoid were 7.6 seconds and 7.1 seconds for the uni-manual and bi-manual approaches respectively, whereas, 7.1 seconds and 6.2 seconds respectively for the large cylinder across both manipulation approaches. This can be attributed to nature of the task, range of motion, and trial duration. We believe that a relatively more complex tasks such as multi-component assembly would have improved the current results resulting in a statistical significance. However, despite of not having a statistical significance, it can be observed that the median completion times for the bi-manual input mode are shorter than the uni-manual mode across both the 3D shapes and their size variants.

Based on the user evaluation results for the Conventional non-immersive VR setup, the bi-manual manipulation showed better overall user-performance in terms of alignment error and task completion time for the spatial manipulation tasks. Based on similar hypotheses, the following section analyzes data for CMWS.

**CMWS: Pairwise Uni-manual vs. Bi-manual Manipulation**

In this sub-section we evaluate the uni-manual and bi-manual input modes for our proposed Clock-Maker’s Work-Space configuration with the intent to have a co-located visual-motor space for allowing precise spatial manipulation of 3D objects. Similar to CVR, the statistical comparisons happen between smaller versions of virtual 3D objects — Cylinder and Trapezoid across both the manipulation approaches and likewise for the larger 3D objects.

**User Performance: Alignment Error**

In order to evaluate the final orientation between the peg and the hole performed by the participants for precise spatial manipulation, we make the following hypotheses:
Null($H_0$): There is no significant difference in the alignment errors for the docking-assembly task in uni-manual and bi-manual input modes.

Alternate($H_a$): There is a significant difference between the alignment errors for the docking-assembly task in uni-manual and bi-manual input modes.

The data samples collected for the uni-manual and bi-manual docking tasks (Figure 7) are evaluated for normal distribution using the Shapiro-Wilk test and normality is disregarded for p-values $< 0.05$. Since all data samples were not found to be normally distributed, we perform statistical analysis using the non-parametric Kruskal-Wallis test for hypothesis testing based on the aforementioned null($H_0$) and alternate($H_a$) hypotheses. We observe a statistically significant difference in the alignment errors for the small and large cylinder with p-values $< 0.01$, favoring the bi-manual input mode. Here, the median values were found to be 0.13 and 0.12, and 0.2 and 0.21 for the uni-manual and bi-manual input modes across the small and large cylinder respectively. In case of the trapezoid, a statistically significant difference was observed only for the smaller version with a p-value of 0.03 having median errors of 0.07 and 0.09 for the uni-manual and bi-manual approaches, whereas, for the larger version, no statistical significance was observed with a p-value of 0.44 having median errors of approximately 0.07 for both uni-manual and bi-manual approaches.

User Performance: Docking-Task Completion Time

In order to evaluate the time taken by the participants to completely dock the peg into the hole, we make the following hypotheses:

Null($H_0$): There is no significant difference in the time taken to completely dock the peg and the hole assembly in uni-manual and bi-manual input modes.

Alternate($H_a$): There is a significant difference in the time taken to completely dock the peg and the hole assembly in uni-manual and bi-manual input modes.

The data samples collected for the uni-manual and bi-manual docking tasks (Figure 8) are evaluated for normal distribution using the Shapiro-Wilk test and normality is disregarded for p-values $< 0.05$. Similar to the error metric, all data samples for task completion time were not found to be normally distributed. Therefore, we perform statistical analysis using the non-parametric Kruskal-Wallis test for hypothesis testing based on the aforementioned null($H_0$) and alternate($H_a$) hypotheses. The docking times across uni-manual and bi-manual input modes for the larger cylinder were found to have a statistically significant difference favoring the bi-manual manipulation. The median completion times were 7.1 seconds and 5.7 seconds for the uni-manual and bi-manual input modes respectively. Whereas, the smaller cylinder had similar medians docking times of 7.32 seconds and 7.5 seconds for the uni-manual and bi-manual input modes respectively. In case of the trapezoid, no statistically significant difference was observed for both size variants having median completion times of 9.1 seconds and 9.4 seconds for the small trapezoid across the uni-manual and bi-manual input modes respectively, whereas, 7.4 seconds and 7.7 seconds respectively for the large cylinder across the uni-manual and bi-manual input modes respectively.

Therefore, based on the user evaluation results for the CMWS setup, the bi-manual input mode showed better user-performance in terms of alignment error and task completion time for the spatial manipulation task. The user preference was also found to be inclining towards the bi-manual approach for better manipulation accuracy, control, and precision during the spatial tasks. We proceed further with evaluating the bi-manual approach for both the setup configurations — CVR and CMWS, so as to comparatively analyze them for user performance in order to understand if our proposition for a co-located visuo-motor space for precise 3D manipulation performs better.

Bi-manual Manipulation: CVR vs. CMWS

In our comparison between the uni-manual and bi-manual input modes for each of our setup configurations, each data sample was evaluated to be non-normal using the Shapiro-Wilk test. Therefore, we only compare the bi-manual user data from CVR and CMWS using the non-parametric Kruskal-Wallis test for hypothesis testing based on the null($H_0$) and alternate($H_a$) hypotheses stated for each of the following user performance evaluation metrics. The comparisons happen between smaller versions of virtual 3D objects — Cylinder and Trapezoid for the bi-manual input mode and likewise for the larger 3D objects.

User Performance: Alignment Error

In order to evaluate the final orientation between the peg and the hole performed by the participants for precise spatial manipulation, we make the following hypotheses:

Null($H_0$): There is no significant difference in the alignment errors for the docking-assembly task in the bi-manual input modes across the two setup configurations.

Alternate($H_a$): There is a significant difference between the alignment errors for docking-assembly task in the bi-manual input modes across the two setup configurations.

No statistically significant difference was observed in the alignment errors (Fig. 9) across both shapes and their individual sizes for the bi-manual input mode. The p-values observed for small and large cylinders were 0.54 and 0.33 respectively.
with approximate median errors of 0.13 for both shapes and their size variants. It is also observed that the error data for the CVR configuration has a comparatively larger spread than CMWS. This is can be attributed to the inconsistencies in user manipulation action discussed in detail in further sections.

**User Performance: Docking-Task Completion Time**

In order to evaluate the time taken by the participants to completely dock the peg into the hole, we make the following hypotheses:

**Null($H_0$):** There is no significant difference in the time taken to completely dock the peg and the hole assembly in the bi-manual input modes across the two setup configurations.

**Alternate($H_a$):** There is a significant difference in the time taken to completely dock the peg and the hole assembly in the bi-manual input modes across the two setup configurations.

Statistically significant difference (Fig. 10) for docking time is observed for the trapezoid with p-values of 0.01 and 0.04 for the small and large size variants respectively favoring the **CVR setup**. The median times for the small trapezoid are 7.1 seconds and 9.4 seconds for CVR and CMWS respectively, whereas 6.3 seconds 7.7 seconds respectively for the large trapezoid. Interestingly, the docking tasks performed using the CMWS setup for orienting the Trapezoid were relatively lengthier compared to CVR. No statistically significant difference was observed for the cylinder having p-values of 0.19 and 0.9 across small and large size variants respectively for both setup configurations. The median times for the small cylinder are 6.9 seconds and 7.5 seconds for CVR and CMWS respectively, whereas 5.6 seconds 5.7 seconds respectively for the large cylinder.

**User Feedback & Observations**

We collected an open-ended user feedback post completion of trials for each interaction approach — Uni-manual and Bi-manual per setup configuration — CVR and CMWS. Across the 1056 trials recorded for all 22 participants, an overall positive feedback was received towards bi-manual approach for both the setup configurations with few exceptions favoring the uni-manual approach for the non-immersive CVR. Most users expressed comfort with the overall idea of performing small-space precise virtual object manipulation tasks in close proximity to their body (as offered by CMWS). We discuss some insightful feedback in conjunction with our own observations during the tasks.

**Setup Configuration**

Participants found CMWS to be more intuitive. One user from the CMWS study group intuitively mentioned, “It felt intuitive to control the objects behind the screen and was able to manipulate as I would do in the absence of the screen.” Whereas for CVR, most users confirmed a “lack of steadiness” while spatially controlling the virtual 3D shapes for the docking task. Qualitative feedback (Fig. 11) received from the NASA-TLX ratings show a significant inclination towards the CMWS setup across all comparison metrics. The median TLX ratings (lower score is better) for CVR and CMWS respectively were (A) Mental Demand - 10 and 7, (B) Physical Demand - 7 and 6, (C) Temporal Demand - 10 and 6, (D) Overall Task Performance - 10.5 and 8, (E) Physical Effort - 10 and 5, and (F) Frustration - 5 and 4. Subsequently, when prompted at the end of a study session, the participants experimenting in the CVR setup conveyed that “their spatial actions were a consequence of the configuration” and “in the physical world, they would be doing the docking task rather comfortably at a position relatively closer to their body”. Also, the same is echoed in the accuracy and precision scores (Fig. 12) rated by the participants based on their experiences. CMWS has a relatively higher median score of 8.0 and 7.0 for accuracy and precision respectively. Thus, the user-reviewed NASA-TLX...
ratings, accuracy and precision scores, and open-ended comments reflect the actual user experience during the docking task for each setup configuration. Hence, it is safe to assume that the Clock-Maker’s Work-Space facilitates a more intuitive, comfortable, adaptable, and efficient way of performing precise spatial manipulation actions.

**Interaction Approach**

Each participant performed the study tasks both with a uni-manual and a bi-manual approach for a given study configuration. Humans inherently are bi-manual creatures, and this helps us with the dexterity to efficiently interact with objects in our proximal space. In general, participants using for both study setups struggled with the uni-manual approach. For CVR, few participants preferred the uni-manual approach since they would “have to do twice the mental processing for controlling the peg and detecting the hole” bi-manually due to a lack of “steadiness” of their hands. Thus, controlling only one shape seemed easier and relatively less frustrating, helping them focus better on the task. As a matter of fact, for the bi-manual approach, most participants preferred moving the hand that controlled the peg while keeping the other one stationary (mostly resting on a flat surface)(Fig. 13 so as to fix the position of the hole in the virtual space as well as increase to steadiness of the hands.

**Object Shape & Size**

Our experiment included two shapes - cylinder and trapezoid further scaled to sizes - small and large per shape. We hypothesized that: (a) it will be difficult to dock the trapezoid relative to the cylinder due to the shape geometry, and (b) docking the larger shapes would be relatively easier due to better control. First, the participants did find the trapezoid less “intuitive” to manipulate as compared to the cylinder. Owing to its lack of rotational symmetry, they spent relatively more time in docking the trapezoid peg and hole pair. Also, the user performance results (Figs. 6,8) corroborate with the same. Regarding shape sizes, participants found it relatively easier to dock the larger shapes as it took less effort to identify the edges of the hole and the larger scale was visually more perceptible to map the peg edges to that of the hole.

**Fatigue**

One of the primary reasons for poor adaptability of SUIs is fatigue. One of the common examples of fatigue is the Gorilla Arm Syndrome [42, 1, 69] that characterizes physical exertion caused due to prolonged suspension of user hands in mid-air to perform spatial actions. While we did not conduct a quantitative assessment similar to Consumed Endurance [36, 37], the qualitative feedback collected from the participants provided interesting insight on the role of fatigue for spatial manipulation across both CVR and CMWS. While the duration per trial was short enough to cause any hand-fatigue, users in the CVR group expressed discomfort about the physical stance for performing the docking task which resulted in unsteady hand movements. Also, there arose a need for one of the hands to be rested on a flat surface while performing spatial manipulation. However, for the CMWS setup, the facilitation of a co-located visuo-motor space helped mitigate any physical exertion, fatigue, or discomfort during the docking session. This is because the users were performing a virtual task with a spatial intent of doing the same in the physical world.

**User Experience**

Our participation pool had 40% participants with and remaining 60% without prior experience in VR/AR systems. While the ones with prior experience in VR/AR shared positive feedback regarding the novelty of the systems, it was also elicited that mid-air fatigue and spatial reasoning was a prevailing issue with existing VR/AR approaches. This could be observed in the non-immersive Conventional VR (CVR) system. However, this did not affect the user experience for the Clock-Maker’s Work Space (CMWS), as most participants adapted to the behind-the-screen interaction approach within 1-2 test trials and were comfortable while performing the actual trials as well. In case of CMWS, one participant stated, “the interactions felt obvious and I could perform the task with my eyes closed.” Few participants did express the need for a tactile feedback to simulate the intersection of virtual 3D shapes during precise alignment. At its core, a co-located peripersonal space was reflected in the way users adjusted their arm and hand posture for performing the spatial object manipulation tasks. Hence, there is fundamental need for exploring and characterizing small-space interaction design patterns and strategies for high-precision manipulation tasks.

**DESIGN IMPLICATIONS**

**When and How Does Interaction Configuration Matter?**

The statistical results comparing the alignment error and completion time for CMWS and CVR were unexpected. This is
evident from the fact that while there was a clear user preference in favor of CMWS, the CVR configuration performed better in terms of task completion time for bi-manual input. Furthermore, the alignment error was comparable across the two interface configurations. This was clearly an unexpected result, particularly when users clearly expressed higher comfort and lower load in case of CMWS. We believe that CMWS, while better positioned for supporting tactile control and proprioception, was limited because the user did not directly see their hand manipulating the object. We believe that a truly see-through augmented/reality interface (such as the ones shown in MixFab [74] and HoloDesk [35]) may help improve accuracy and reduce completion time for CMWS.

Technological Challenges
Four technological factors are crucial for peripersonal manipulations: (a) controllers, (b) object tracking, (c) tactile and visual feedback methods. (a) While hand-held controllers are fairly common and robust, other technological aspects need careful consideration. This is especially true since current AR headsets are primarily designed for interaction at a distance. (b) A major challenge that we faced was the lack of guidelines for implementing a close-range motion capture system for tracking high-precision interactions. This was not surprising since motion capture is predominantly used in room-sized environments. On the other hand, we also avoided the use of commodity cameras such as the Leap Motion [78], Kinect [76] and RealSense [75] since they do not offer as precise and smooth tracking capabilities or offer limited features in terms of tracking hand-held objects with high tracking fidelity. Even though there has been a large body of work dedicated to hand and object tracking in the last few years, more development and standardization is needed for making peripersonal interactions commonplace. (c) We also posit that the the broader area of haptics and tactile feedback are currently the most rich areas for further development in this regard. Our studies shed light on the lack of a tangible feedback from the hand-held controllers to indicate that the docking was successful. While not a strong feedback, we believe it will play a significant role for high-precision interactions in intimate spaces where the sense of touch overshadows visual feedback in many cases [20].

Interaction Design
The important aspect of performing any interaction lies in perceiving the immediate surroundings of the human body. Moreover, human anatomy, object geometry and size, and input modes have a significant effect on the interaction space, which is reflected in the action performed. Profitt et al. [62] discuss how the visual perception of distances and sizes is affected by growing expertise making it easier to target the ball. Similar thoughts are shared in the context of tool handling distance and its action-specific size perception owing to the type of task performed [81]. Taking cues from these works, it is important to re-think existing spatial interactions — pointing, ray-casting, HOMER, etc., and the economy of these actions expended during spatial interactions. It is very important that we leverage on visual perception for enabling a co-located visuo-motor space that enhances action-specific perception for close and distant spatial interactions.

CONCLUSION
While our primary goal was to understand spatial interactions in the user’s peripersonal space, this work demonstrates that there are fundamental interaction design challenges that need to be addressed before generalizing such interactions for high-precision spatial manipulation tasks. The current work specifically explored the role of uni-manual and bi-manual inputs for object manipulation in two interaction configurations — the clock-maker’s works-space and conventional non-immersive VR. Our study reveals two important aspect of close space object manipulation. First, the geometry and the size of the shape being manipulated affects user performance and comfort. Second, although some of our tests showed that CVR performed better in some specific cases, our qualitative results show that CMWS is preferred by most users. These two observations indicate a richer set of research directions to investigate on high-precision object manipulation in the clock-maker’s work-space.

REFERENCES


337–339. DOI: http://dx.doi.org/10.1177/001872998862800399


[64] Benjamin Schmidt. 2002. When are two hands better than one?: a study of bimanual interaction. (2002).

[65] Andrea Serino, Jean-Paul Noel, Giulia Galli, Elisa Canzoneri, Patrick Marmaroli, Hervé Lissek, and Olaf Blanke. 2015. Body part-centered and full body-centered peripersonal space representations. Scientific Reports 5 (12 2015), 18603. DOI: http://dx.doi.org/10.1038/srep18603


